



the globus alliance
www.globus.org

Identifiers, Resources, EPRs, and Missing Links

OSG - Middleware Security Group Meeting

Mon-Tue, June 5-6, 2006, SLAC, Stanford, CA

Frank Siebenlist

(Argonne National Laboratory / University of Chicago)

franks@mcs.anl.gov - <http://www.globus.org/>





the globus alliance

www.globusalliance.org

W3C WS-Addressing's Endpoint References (EPR)

- “A Web service endpoint is a (referenceable) entity, processor, or resource to which Web service messages can be addressed.”
- “Endpoint references convey the information needed to address a Web service endpoint.”
- “Endpoint Reference Comparison. This specification provides no concept of endpoint identity and therefore does not provide any mechanism to determine equality or inequality of EPRs and does not specify the consequences of their equality or inequality. However, note that it is possible for other specifications to provide a comparison function that is applicable within a limited scope.”



Issues?

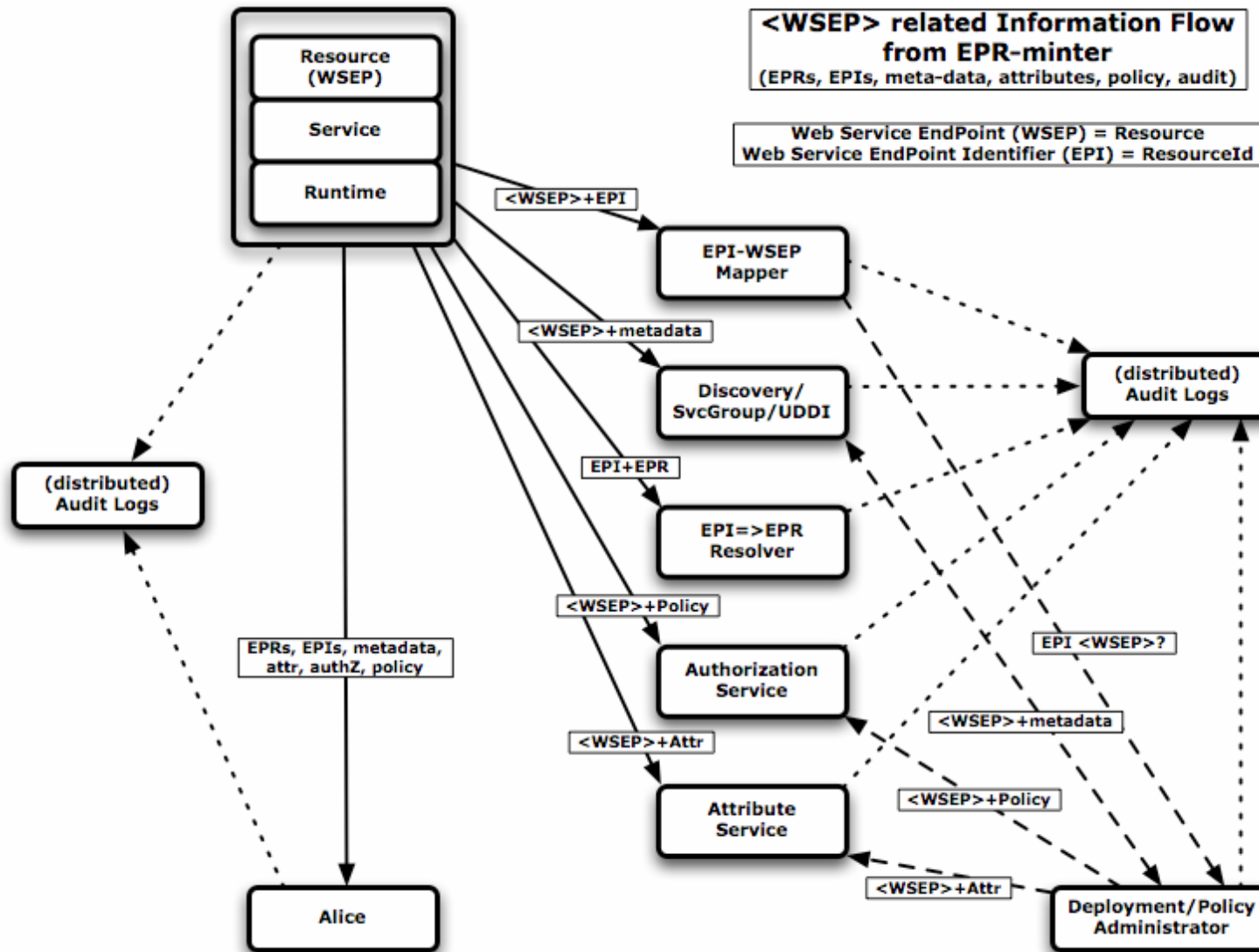
- No way to compare EPRs...
 - ◆ How to associate policy/audit with them
 - ◆ How to “know” whether two EPRs refer to same resource
- Where does the EPR point to tomorrow?
 - ◆ Today it refers to your bank account...
 - ◆ Tomorrow it may refer to yours...
 - ◆ (one of us will be unhappy...)



Resource Identifier Use Case

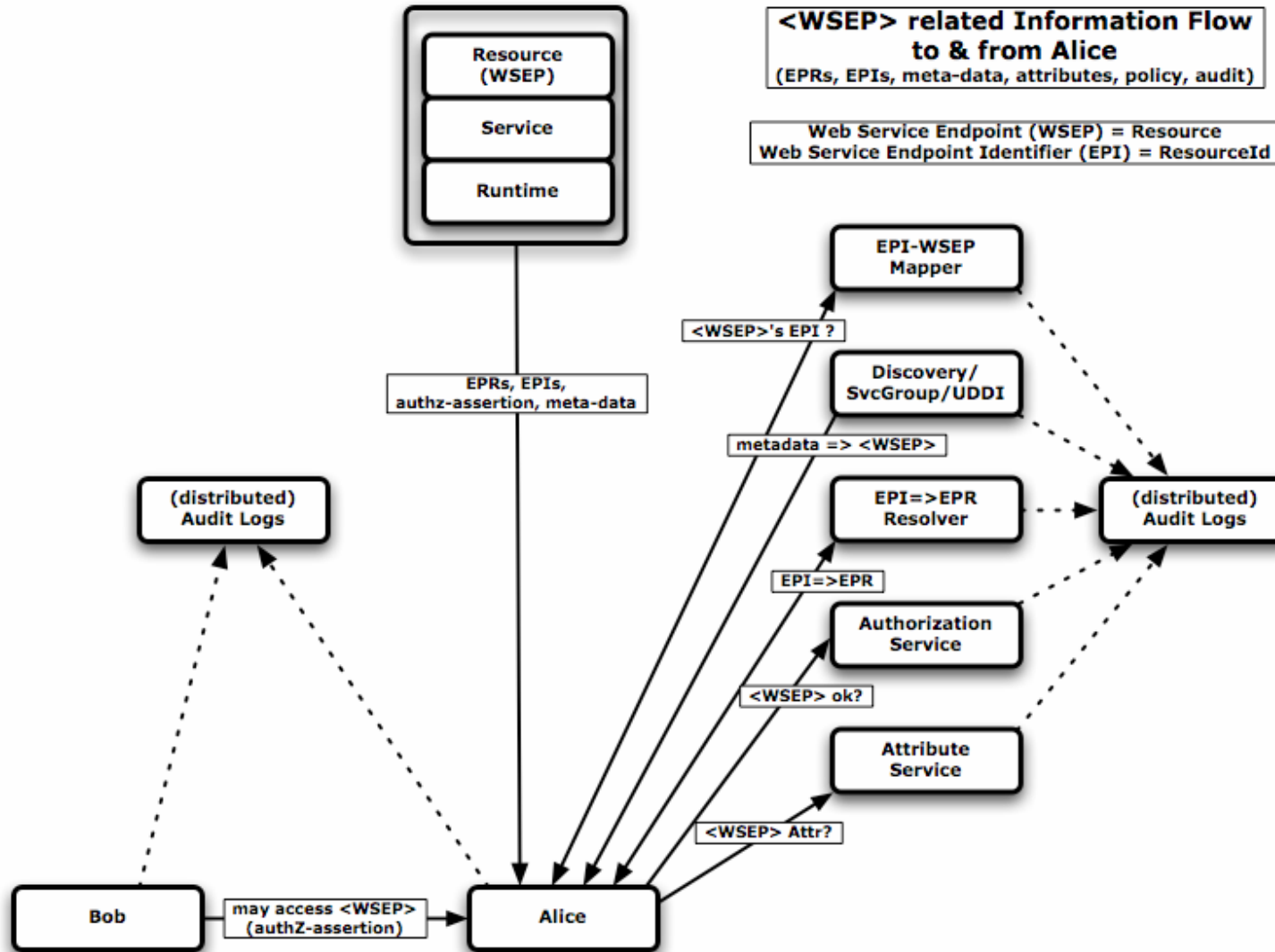
- **Resource Mobility.**
- **Assertion Target.**
- **Resource Attributes**
- **Resource Reference Consistency**
- **Resource Metadata Caching**
- **Audit Label**

EPR Minter & Endpoint Identifiers



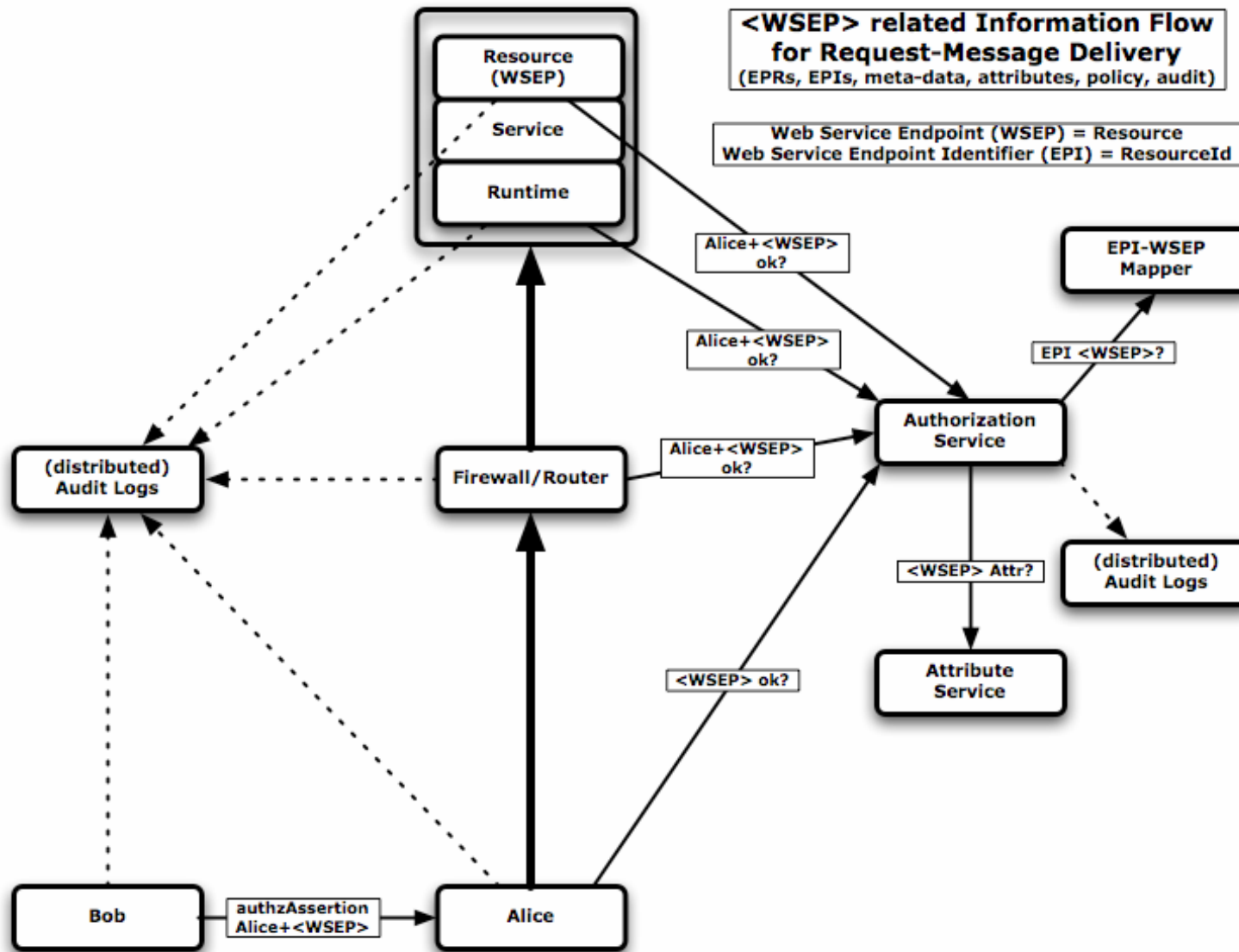


EPR & Identifier Consumer





EPR, EPI and Message





Resource Identifier requirements

◆ *required*

- 1. *Consistency with current tooling*
- 2. *Unambiguous referencing*
- 3. *Client side resource-equality testing*
- 4. *A resource identifier in every message.*
- 5. *EPR resolution*

◆ *desirable*

- 6. *Works with current/existing tooling*
- 7. *Consistency with W3C architecture*
- 8. *Unique address*

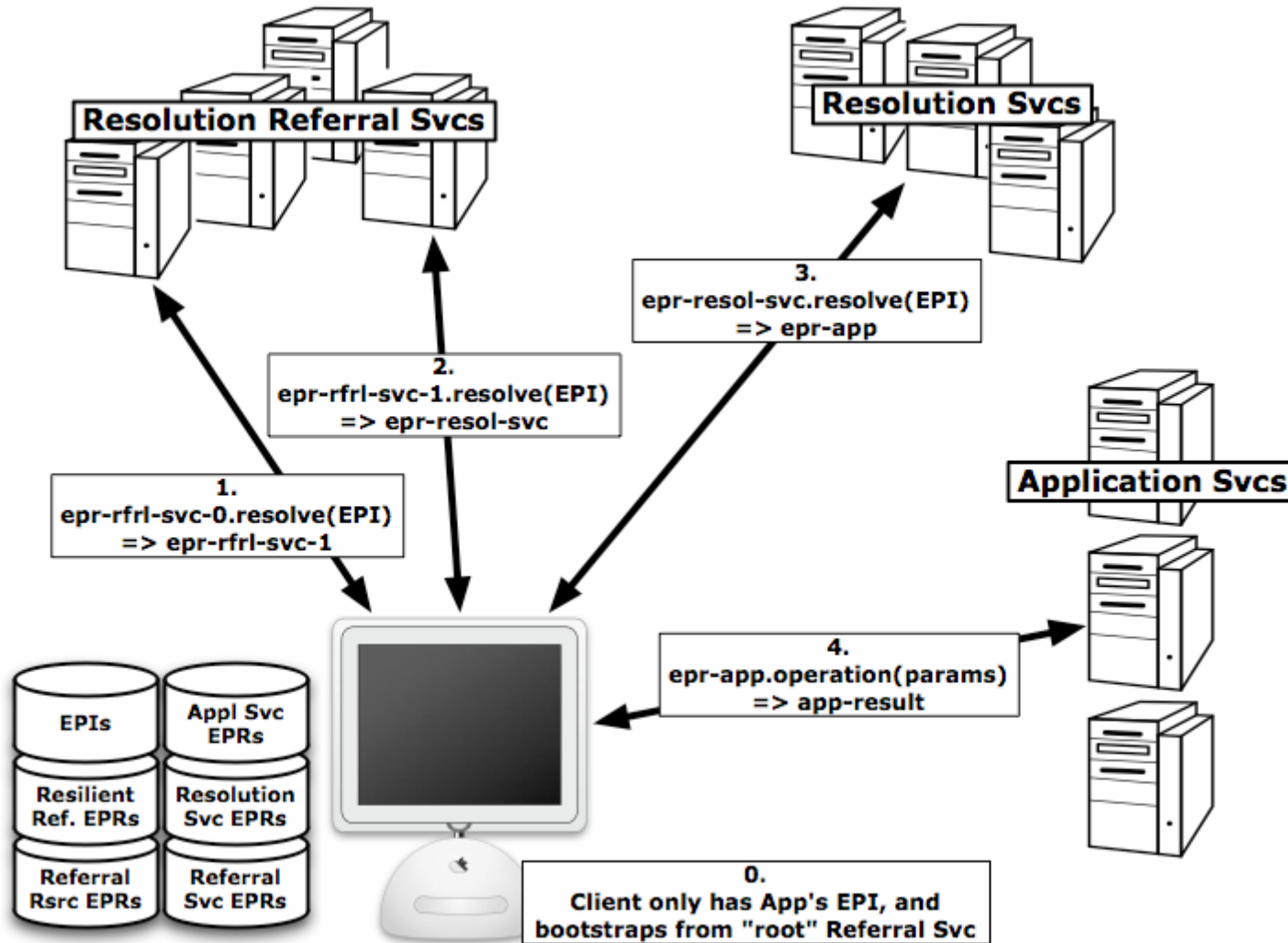


GGF WS-Naming

- **Specifications:**
 - ◆ Web Service Endpoint Identification and Resolution: Use Cases and Requirements
 - ◆ Unambiguous Web Service Endpoint Profile
 - ◆ Web Service Endpoint Address Identifier Profile
 - ◆ Web Service Endpoint Name Specification
 - ◆ Endpoint Reference Resolution Specification



EPR Resolution Svcs (from EPI)





caBIG

- Cancer Grid project by NCI/NIH
 - ◆ The **cancer Biomedical Informatics Grid**, or **caBIGI**, is a voluntary network or grid connecting individuals and institutions to enable the sharing of data and tools, creating a World Wide Web of cancer research. The goal is to speed the delivery of innovative approaches for the prevention and treatment of cancer. The infrastructure and tools created by caBIGI also have broad utility outside the cancer community. caBIGI is being developed under the leadership of the [National Cancer Institute's Center for Bioinformatics](#).
 - ◆ BIG project: Over 800 people from more than 80 organizations are working collaboratively on over 70 projects in a three-year pilot project.
 - ◆ <https://cabig.nci.nih.gov/>



Identifier Services Framework

- Identifier
 - ◆ “Naming” of individual Data-Objects
 - ◆ Globally Unique Name for each Data-Object
- Services
 - ◆ Create/modify/delete name-object bindings
 - ◆ Resolve name to data-object
- Framework
 - ◆ Provide for Trust Fabric => Binding Integrity
 - ◆ Policy-driven Administration => Curator Model
 - ◆ Fully Integrated with caGrid’s Architecture and Implementation



the globus alliance

www.globus.org

Why (Standardized) Resource Identifiers?

- Efficiency
 - ◆ Passing by reference vs by value (Data-Object can be many Mbytes)
 - ◆ Data-Object Equality test through String comparison (inequality test is no requirement...)
- Consistency
 - ◆ Standardized way of referencing objects
 - ◆ Standard identifier => data-object resolution mechanism
 - ◆ Meta-data binding to standard object reference
 - ◆ Well-known primary/foreign key for (distributed) JOINS
 - ◆ Name for policy expression for data-object access
 - ◆ Name for audit entries about data-object related activities
 - ◆ ...
 - ◆ Possible correlation of all of the above...



Data-Object Identifier Properties

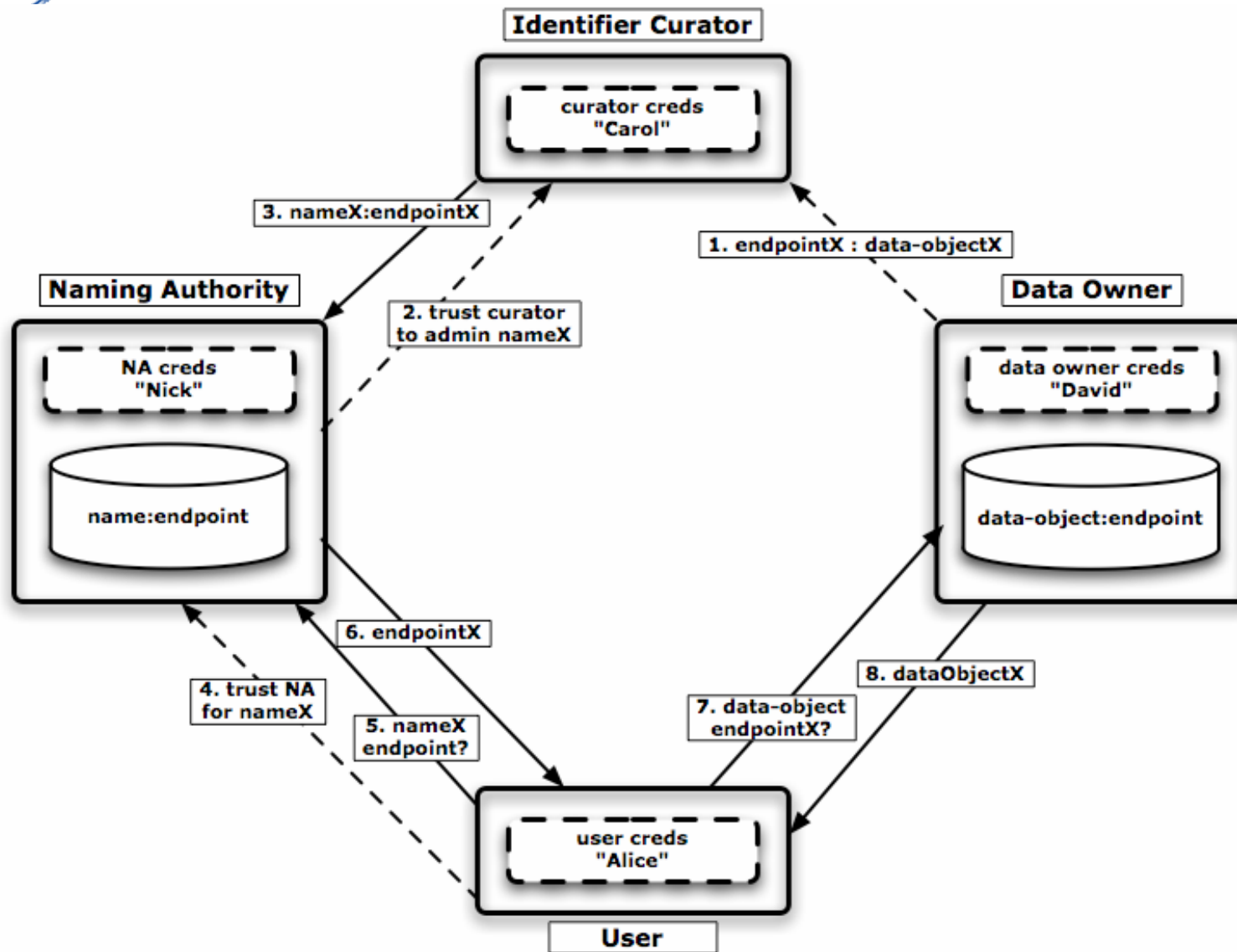
- Identifier is a String
- Identifier is a forever globally unique name for single Data-Object
- Identifier can be (globally) resolved to associated Data-Object
- Data-Objects are immutable, almost immutable or mutable...

- Identifier value “meaningless” opaque string for consumer
- Resolution information embedded in Identifier Name
 - ◆ Only meaningful for resolution service related components

- Identifier is a Universal Resource Identifier (URI)



Identifier Usage Model





Naming Authority, Identifier Curator, Data Owner and Identifier User

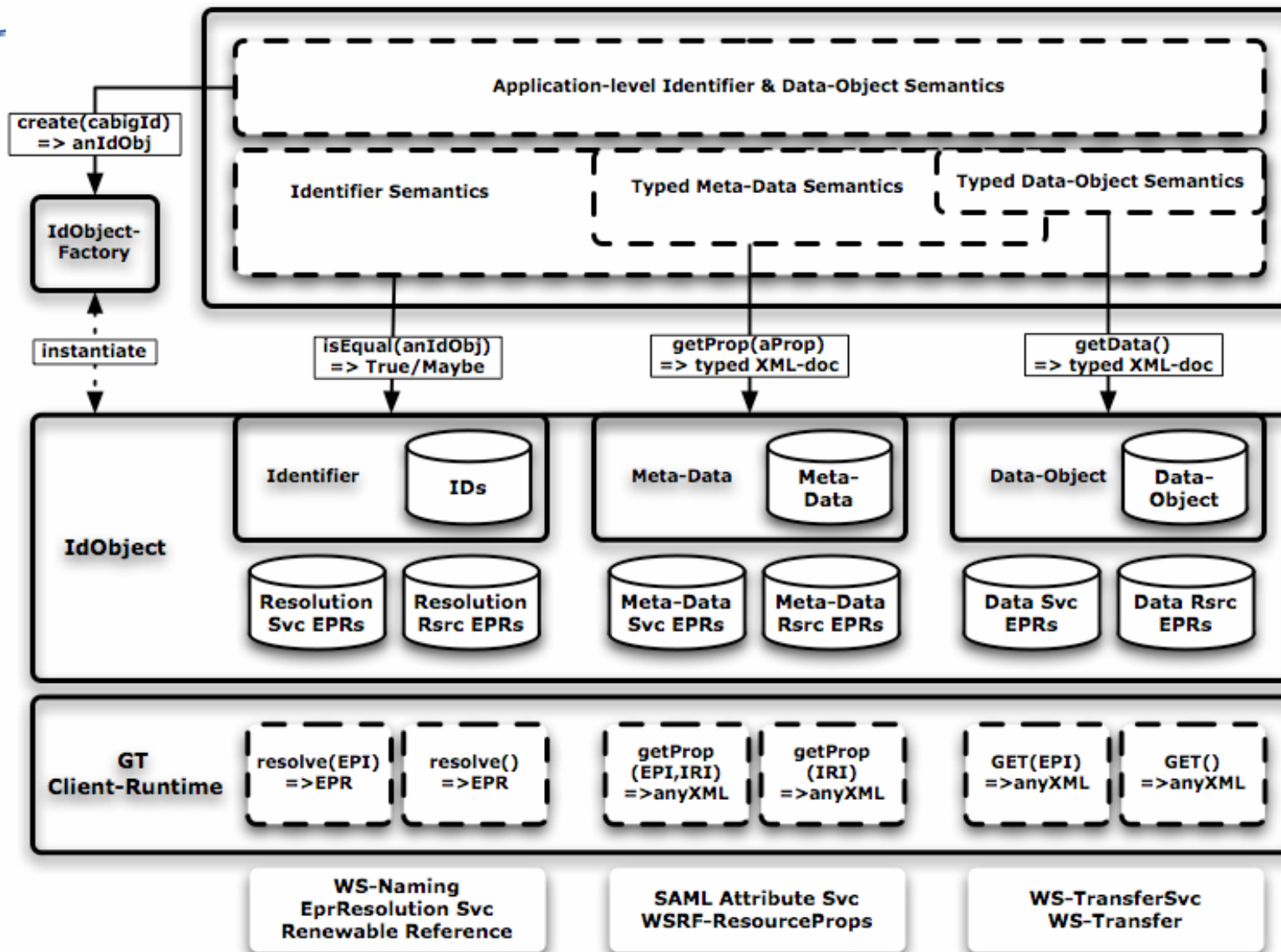
- Naming Authority (NA)
 - ◆ Guards integrity of identifier namespace & bindings
 - ◆ Maintains identifier to data-object's endpoint mapping
- Identifier Curator/Administrator
 - ◆ Understands semantics/access of data owner's objects
 - ◆ Trusted by NA to administer binding for certain identifiers
 - ◆ Administers identifier to data-object's endpoint binding
- Data Owner
 - ◆ Provides access to data-objects through "endpoint-references"
- Identifier User/Consumer
 - ◆ Trusts an NA for certain identifier bindings
 - ◆ Uses 2-step resolution to obtain data-object (identifier => endpoint => data-object)
 - ◆ (In-)Directly trusts Data Owner for data-object integrity

Conclusion

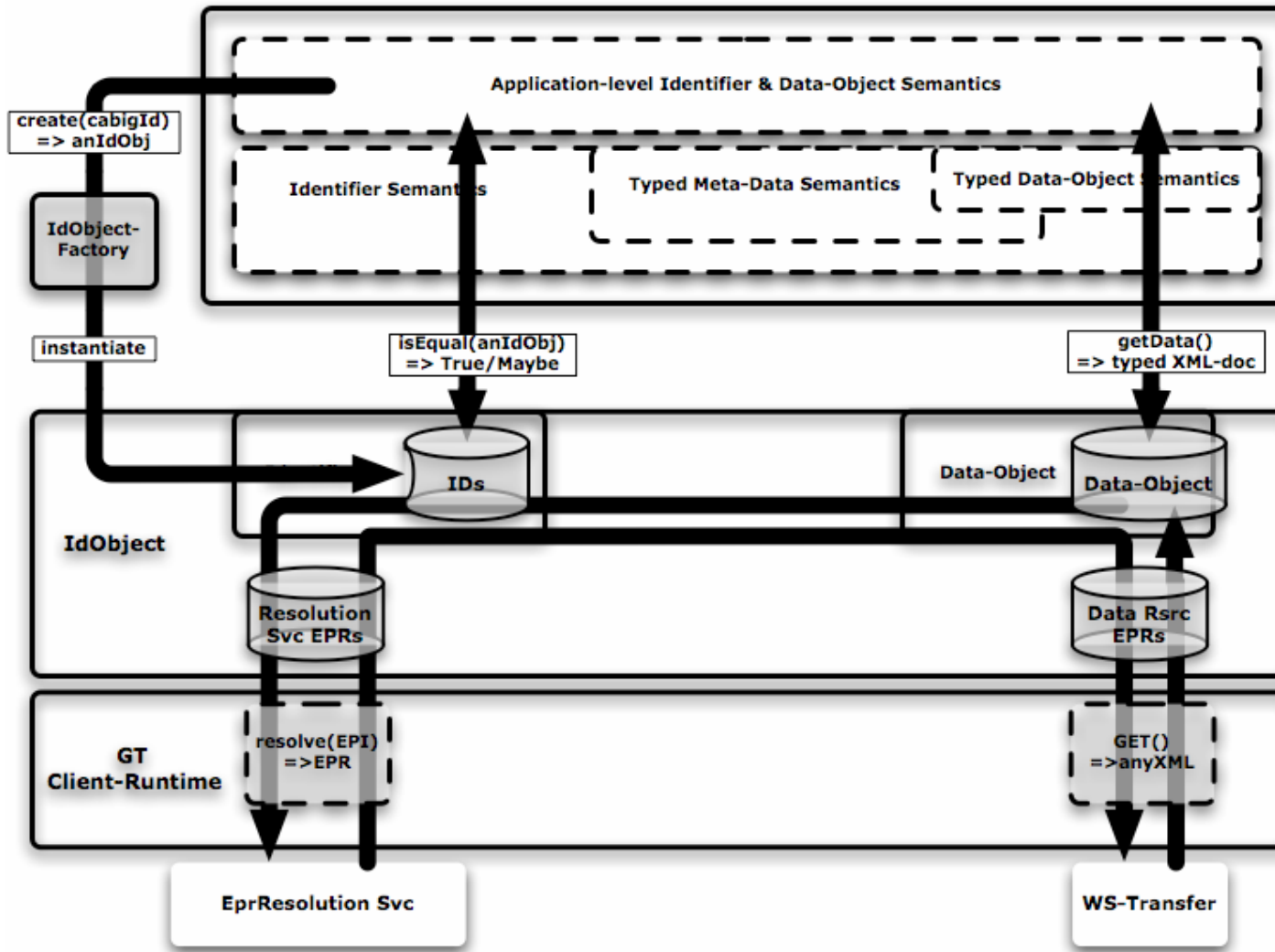
- Current WS-Addressing not good enough!
- Need for profiles to require unambiguous use of EPRs
- Need standardize identifier usage for policy/audit !!!
- Need identifier services framework to provide the trust fabric for the bindings



Identifier Consumer



Identifier Consumer First Step



Identifier & Data-Service

